

The production of L3 stop-initial words by Spanish/English bilinguals

Kyle Parrish¹

¹ Rutgers University

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Kyle Parrish,
Department of Spanish and Portuguese, Rutgers University, New Brunswick, United States.
E-mail: kyle.parrish@rutgers.edu

Abstract

The present study examined the production of L3 French words by Spanish-English bilinguals who had no prior knowledge of the L3. Using a shadowing task, 39 Spanish L1/English L2 and 30 Spanish monolingual speakers produced 26 tokens of word-initial voiceless plosive consonants in French, Spanish, and English (15 Spanish and French tokens for the monolingual group). Importantly, Spanish and French plosives are typically produced with short-lag VOT, whereas English plosives are produced with a long-lag VOT and aspiration. The results of generalized mixed effects models revealed that L3 French productions by the bilingual group fell between L1 and L2 values, and provided evidence that the bilingual group experienced influence from their both their L1, Spanish and their L2, English, while the monolingual group produced the French /p/ as a Spanish-like short-lag /p/. These findings provide support for the Linguistic Proximity Model (Westergaard et al., 2017), and highlight the need in L3 research to employ larger sample sizes, as well as a need to determine whether results of previous studies can tease apart acquisition from cross-linguistic influence.

Keywords: third language acquisition, voice-onset time, cross-linguistic influence, multilingualism

Word count: X

The production of L3 stop-initial words by Spanish/English bilinguals

Introduction

Third language (L3) acquisition can be defined as the process of learning or acquiring an additional language by an individual who already speaks two languages. Recent models of L3 acquisition have suggested that this process is not simply an additional instance of second language (L2) learning, but rather that it merits its own field of research. That is, the field of L3 acquisition has generally moved towards a perspective that, unlike the L2 learner whose new language is bound to be influenced by their mother tongue, the L3 learner has two linguistic systems from which the L3 can be influenced. In general, the study of L3 acquisition provides insights into not only how bilinguals learn a new language, but the nature of their previously acquired languages. Models in third language acquisition debate the role of previously acquired languages during the acquisition of a third. Various models of L3 acquisition endeavor to predict the process by which bilinguals acquire a new language. Some of these models suggest that only one language wholly influences the acquisition of a third, rather than gradient and simultaneous influence of two languages. The L2 Status Factor Model (Bardel & Falk, 2007, 2012) predicts that the second learned language will block access to the L1, and that the L3 will be influenced solely by the L2. On the other hand, the Typological Primacy Model (TPM; Rothman (2010), Rothman (2011), Rothman (2013), and Rothman (2015)) posits that either the L1 or the L2, but not both, will influence the L3, and that the choice of which of these two languages influences the L3 is determined based on perceived typological similarity on the basis of L3 input. Other models suggest that whole language influence does not occur, and that gradient effects of both languages is possible such as the Scalpel Model (Slabakova, 2017) and the Linguistic Proximity Model (LPM; Westergaard, Mitrofanova, Mykhaylyk, and Rodina (2017)). The present study aimed to provide evidence for the predictions of these L3 models by measuring the voice-onset time (VOT) of French words produced by Spanish-English bilinguals who do not speak French. VOT was chosen as a feature of interest due to the cross-linguistic differences between the

chosen languages. French and Spanish are similar in their use of VOT to distinguish stop consonants since both languages are so-called true-voicing languages (Lisker & Abramson, 1964). English, on the other hand, is an aspirating language which distinguishes stop consonants by way of long and short lag VOT, where phonemically voiced stops are often phonetically voiceless.

In order to inform the predictions of L3 models, the present study examines the relative VOT of French stops by Spanish-English bilinguals in order to examine whether speakers produce voiceless stop consonants using aspiration, despite the Spanish-like phonetic qualities of the input. In addition to the experimental design, the present study also was intended to include a statistically powered sample, which have been scarce in the L3 literature to date. Low sample size and low power are associated with an increased risk of sampling error, and the commitment of type 1 and type 2 errors (Brysbaert, 2020). In order to obtain a larger sample, the present study recruited absolute L3 beginners. Recent studies in the acquisition of L3 VOT have involved a range of L3 proficiencies, in which it is possible that the results of studies could also be explained by acquisition of language structures by more proficient learners, rather than cross-linguistic influence.

By using bilinguals who are encountering an L3 for the first time, the chances of acquisition as a possible confound are greatly reduced. By introducing these changes in the typical L3 methodology, the present study provides updated and higher statistically powered evidence to examine the nature of cross-linguistic influence on third language production at first exposure.

Literature Review

Models of L3 acquisition fall into two major categories: full transfer models and hybrid transfer models. The most prominent full transfer models include the Typological Primacy Model (TPM; Rothman, 2010, 2011, 2013, 2015), and the L2 Status Factor Model (L2SF; Bardel & Falk, 2007). Both of these models predict that one language will serve as the sole

basis of influence for the third language. In the L2SF, the L2 is predicted to influence the L3 by default due to the cognitive similarity of late-learned languages and is based on the Declarative/Procedural Model (Paradis, 2009; Ullman, 2004). The TPM, on the other hand, predicts that either the L1 or the L2 influence the L3, but not both. In this view, the choice of L1 or L2 is individually determined on a subconscious level after exposure to input. The input is parsed for cues in a hierarchical manner in the TPM, in which the lexicon is theorized to be the primary cue attended to by the L3 learner, followed by phonological cues, then by morphosyntax and finally by syntax (Rothman, 2015). Hybrid transfer models include the Scalpel model (Slabakova, 2017), and the Linguistic Proximity Model (Westergaard, Mitrofanova, Mykhaylyk, & Rodina, 2017) and the Cumulative Enhancement Model (Flynn, Foley, & Vinnitskaya, 2004). These models predict that both the L1 and the L2 may influence the L3, but do not generate specific predictions regarding which of the two choices influence L3 learning and the conditions necessary for that influence to occur. Previous studies in L3 acquisition have yielded mixed results and have examined various linguistic domains. In a recent systematic review, Puig-Mayenco, González Alonso, and Rothman (2020) reviewed 71 empirical studies (largely focused on L3 syntax) obtained from journals, books, conference presentations and unpublished dissertations and theses. The authors were examining the results of the studies in order to gather cumulative evidence for the predictions of L3 models. The majority of the studies were coded to support the predictions of the TPM, since influence was found to come from either the L1 or the L2, or the L2 status factor, when influence was found to have come solely from the L2. The systematic review also found evidence for hybrid models, though this was evidenced in fewer total studies. Studies in L3 phonology have largely not supported the predictions of the TPM to date, while they have provided mixed evidence for the L2 status factor. In an influential case study, Williams and Hammarberg (1998) studied the language of a highly proficient bilingual (L1 British English L2 German) adult female who had recently moved to Sweden. A recording of the subject speaking L3 Swedish was recorded and presented to

native speakers of Swedish, who were not told the linguistic background of the person in the recording. The task was for the Swedish informants to guess the linguistic origin of the speaker in the recording. The results revealed that the Swedish speaking informants interpreted the speech of the speaker in the recording as German-accented Swedish, the speakers' L2. The results of this study corroborated the so-called "foreign language effect" first suggested by (Meisel, 1983), the idea being that L3 learners attempt to sound unlike a speaker of their L1 in the productions of a new language. However, a follow up of the same task with the same speaker 6 months later produced the opposite result. The informants rated the speaker as having English-accented Swedish. Further studies in L3 phonology yielded mixed results, with some studies reporting L2 status effects (Gut, 2010; Llama, Cardoso, & Collins, 2010; Wrembel, 2010), while others concluded that there was mixed influence of both the L1 and the L2 (Llama & Cardoso, 2018; Wrembel, 2014).

The mixed results in the literature may have a number of causes, including low sample size and the choice of analysis used. In earlier studies at the beginning of the 2010's, ANOVA was a common choice of analysis. For example, in his seminal paper, Rothman (2011) examined two L3 groups using a one-way ANOVA. The results of the analysis did not reveal a main effect of group, which the author took as evidence for identical performance of the groups. Two potential issues with this conclusion is that, firstly, the sample size was quite low ($n = 12$ and $n = 15$) per group, and secondly a test of equivalence (Lakens, 2017) was not used in order to determine comparability across groups. In other words, the lack of a main effect in a one-way ANOVA does not in itself constitute evidence for equivalence. Rather, this result suggests that the ANOVA provides inconclusive evidence for equivalent between-subjects performance on the experimental task. This analytical strategy has been common in studies of L3 syntax, and is evident upon closer examination of the systematic review of Puig-Mayenco and colleagues (2020). Many of the studies included in the systematic review had small sample sizes, were not coded in the systematic review consistently with authors' conclusions in the study itself, and none used a test of equivalence

to conclude that groups performed similarly. Instead, many studies used the lack of a main effect in an ANOVA, and sometimes used the lack of a significant p-value on post-hoc pairwise comparisons.

An additional reason that mixed results have been reported may also be due to the variable proficiency levels in the L3. Some studies examined beginners, while others examined L3 learners of intermediate or advanced proficiency. The present study addresses this range in L3 proficiency by presenting words to bilinguals in a language that they do not know. This way, bilingual participants can be thought of as absolute L3 beginners, and a starting point for L3 phonological acquisition can be determined.

The present study

The present study aims to address the potential issues of low sample size and confound of acquisition and facilitative cross-linguistic influence by recruiting a large sample of Spanish-English bilinguals and exposing them to French words. By measuring the relative VOT of L3 French at first exposure, insights can be gained as to how the phonology of the L1 and L2 interact with initial L3 production. In particular, the present study is guided by the following research question: When Spanish-English bilinguals produce French words at first exposure, will their VOT productions be more L1 or L2 like? It was predicted that L3 relative VOT productions would be practically equivalent to L2 VOT and provide evidence for the L2 Status Factor, but not for the Typological Primacy Model. On the other hand, if the results reveal intermediate relative values in which the L3 values fall between L1 and L2 values, hybrid transfer models such as the LPM and Scalpel Model would be supported.

Methods

Participants

A total of 76 participants were recruited for the present study. Two groups of participants were recruited, consisting of an experimental group and a control group. After

exclusion of some participants, the experimental group consisted of 39 Mexican Spanish-English bilinguals and the control group consisted of **30** Spanish monolinguals. Bilinguals were excluded if they did not produce distinct relative VOT in their L1 and their L2 ($n = 37$), or answered ‘yes’ to the question ‘do you speak a third language?’ The experimental group consisted on late bilinguals who acquired their L2 English at an average age of (\bar{X} , $sd = x$) and their average proficiency in English was (\bar{X} , $sd = x$) on self-rated scale of 7.

[Insert table 1]

Materials

Three total tasks were used to elicit VOT production. A shadowing task was used to gather French relative VOT, in which participants heard a recording of French words in isolation and repeated those words. In English and Spanish, elicited production tasks were used. These tasks presented a single stimulus on the screen and participants recorded themselves reading the word aloud.

To gather background information and to measure proficiency, a brief language questionnaire was given to participants. Participants were asked their country of residence and of origin, whether they spoke a third language, whether they spoke a second language, and their self-rated proficiency on a scale from 1 to 7, with 7 being the highest possible score and 1 being the lowest possible score.

The stimuli were words with word initial stops of either one or two syllables, with stress on the initial syllable (see table 2). The French words were produced by a French native speaker and were produced with Spanish-like relative VOT (Mean VOT = 0.06, $sd = 0.04$).

[insert table 2]

Procedure

The elicited production task and shadowing task were programmed in Labvanced, a browser based software program designed for online research. All tasks were given in one session. The order of the languages was counterbalanced and the stimulus order was randomized with each language specific block. The questionnaire was given at the beginning of the tasks. All instructions were given in Spanish.

Statistical Analyses

Sample size was determined by carrying out a power analysis (Brysbaert, 2020), in which a small sample of pilot data was used to determine an appropriate sample for within-subjects tests of equivalence (Lakens, 2017) where the equivalence bounds were $-.4$ and $.4$ standard deviations from the mean, alpha was $.05$, and power was $.8$.

To determine the relative effect for language, the production data were analyzed using generalized linear mixed effects models (Baayen, Davidson, & Bates, 2008) in which relative VOT was modeled as a function of language (French, Spanish, English and consonant (p, t, k). Relative VOT was calculated by dividing the duration in milliseconds of the VOT by the duration of the entire word in milliseconds and converted to z-scores in each language. Relative VOT has been proposed in the literature as an alternative to absolute VOT to account for speech rate (Stölten, Abrahamsson, & Hyttenstam, 2015). For each model, subjects and words (item) were random intercepts, and random slopes were included by participant per language. The residuals were determined to be normal by visual inspection of a Q-Q plot and a residuals versus fits plot. In order to determine the individual contribution of each fixed effect predictor and to derive main effects, nested model comparisons were carried out. These nested model comparisons began with an intercept-only model and added the fixed effect predictors of language and consonant one at a time.

To test for practical equivalence, participant mean relative VOT per language were

utilized to carry out paired tests of equivalence (TOST; (Lakens, 2017)). The equivalence bounds were set to $-.4$ and $.4$ standard deviations from the mean respectively, a small effect size reported in L2 literature ((Plonsky & Oswald, 2014)).

Finally, initial and post-hoc t-tests were used for inclusion and subsetting procedures. The initial t-tests were run per participant to determine whether there was a difference between Spanish and English relative VOT. The post-hoc t-tests were used as a method of subsetting the data into subgroups in which participants showed greater influence of either L1 or L2 on L3 productions.

Results

Tables 3, 4 and 5 list the relative VOT per consonant in each language and table 6 contains the pooled relative VOT across place of articulation per language for the full dataset.

[insert tables 3-6]

Figure 1 plots relative VOT as a function of language. An inspection of the box plot suggests that French relative VOT falls between that of English and Spanish.

[insert figure 1]

Nested model comparisons revealed a main effect for both *language* ($X(2) = 19.97$; $p < .001$) and *consonant* ($X(2) = 28.55$; $p < .001$). No main effect was found for the language x consonant interaction. Overall, the best-fitting model explained 29.80% of the variance observed with fixed effect predictors (marginal R squared), and 54.90% of the variance in the data with the added random effects (conditional R squared). The parameter estimates of the best fitting model suggest, with the effect of consonant held constant and English as the baseline, relative VOT increases by -1.27 standard deviations in the case of Spanish and -0.66 standard deviations in the case of French. These parameter estimates, coupled with the

visual information provided by the boxplots, suggest that, overall, French productions fall between those of Spanish and English. The full output of the model can be seen in table 7.

[insert table 7]

Figure 2 shows the results of the two sample test of equivalence between French and English productions. As one can see, the result of the null hypothesis significance test (a t-test) corroborates the output of the linear model and suggests that English and French relative VOT productions are distinct. As a result, the samples were not determined to be equivalent.

[insert figure 2]

Interim discussion

The results of the GLMM and the TOST do not provide support for the prediction that the L3 would be produced in an L2-like fashion. On the contrary, the results of these analyses show that French relative VOT falls between L1 Spanish and L2 English relative VOT. However, these analyses alone do not make clear whether individual variation is the cause of the group trend, or whether most participants display intermediate relative VOT. In the context of L3 models, the TPM and the L2SF would not expect hybrid values to occur, where models such as the LPM and Scalpel model can account for these results.

In order to better evaluate whether co-activation of L1 and L2 or activation of a single language is occurring, the full dataset was subset into primary L1 and L2 influence post-hoc. Participants were divided into groups based on their individual French-English t.test results. When a difference between French and English was detected ($p < .05$, $n = 16$), participants were placed in the L1-influence subset, where when no difference between French and English was detected ($p > .05$, $n = 23$), they were placed in the L2 influence group. The same statistical tests used for the full dataset, including generalized linear mixed effects models

and tests of equivalence, were used to analyze the data subsets.

Post-hoc results

TOSTs revealed that the L1 Spanish $t(239.88) = -2.458$, $p = 0.00733$ and **L2 English** ($t(309.9) = -2.832$, $p = 0.00246$) productions were produced practically equivalent between the subsets. These results suggest that the participants in the two data subsets produced practically equivalent relative VOT values in both English and Spanish, and it is likely that differences in the L3 relative VOT is not due to a difference in the Spanish and English productions of the subsets in comparison to one another.

L2 subset model. The L2 subset group was determined on the basis of two sample t-tests in which individual participants' French and English relative VOT productions were determined to be inconclusive ($p > .05$). Descriptive pooled VOT across place of articulation for the L2 subset is seen in table 8. The L2 subset displays the same trend as the full data set; the L3 relative VOT productions fall between L1 and L2 values.

[insert table 8]

The L2 subset model is plotted in figure 3. Relative to the full data set model (figure 1), the French values appear to be more heavily influenced by the L2. Like the full data set, nested model comparisons revealed a main effect for both *language* ($X(2) = 18.47$; $p < .001$) and *consonant* ($X(2) = 26.73$; $p < .001$). However, the parameter estimates of the model revealed a smaller effect for language for French -0.34 relative to English than the full dataset model (-0.66), suggesting that the L2 subset had greater influence of English in their L3 French. The effect for Spanish was similar -1.19 (full model -0.66). The full L2 subset model can be seen in table 9.

[insert figure 3]

[insert table 9]

The results of the TOST for the Spanish and French productions of the L2 subset can be seen in figure 4. Both the Test of Equivalence and the Null Hypothesis Significance Test (NHST) were inconclusive, suggesting that French and English productions cannot be concluded to be distinct, nor practically equivalent in this subset.

[insert figure 4]

L1 subset model. The L1 subset group was also determined on the basis of two sample t-tests, but, oppositely from the L2 subset, individual participants' French and English relative VOT productions were determined to be distinct ($p < .05$). Descriptive pooled VOT across place of articulation is seen in table 10. The L1 subset descriptive data reveals that this subset of participants appears to have greater L1 influence on L3 productions than the L2 subset did for L2 influence on L3 productions.

[insert Table 10]

The L1 subset data is plotted in figure 5. Relative to the full data set model (table 7), the French values appear to be closely resemble the L1 values. Nested model comparisons found that the best fitting model was the same as the full data model and the L2 subset model, in which there was a main effect for both *language* ($X(2) = 24.59$; $p < .001$) and *consonant* ($X(2) = 26.56$; $p < .001$). The parameter estimates of the L1 subset model revealed similar effects of language relative to English (French: -1.13, Spanish: -1.39), suggesting that the French and Spanish are similarly distant from English relative VOT values.

[insert figure 5]

Additionally, the parameter estimates of the model revealed a larger effect for language for French -1.13 relative to English than the full dataset model (-0.66). The effect for Spanish was similar -1.39 (full model -1.27). The full L1 subset model can be seen in table 11.

[insert table 11]

The results of the TOST for the Spanish and English productions of the L1 subset can be seen in figure 6. Again, both the Test of Equivalence and the Null Hypothesis Significance Test (NHST) were inconclusive, suggesting that French and Spanish productions cannot be concluded to be distinct, nor practically equivalent in this subset. However, in this case, the absolute mean difference between Spanish and French is smaller than the case of the L2 subset with English and French (.0009 vs .0024), implying that the L3 productions of the L1 subset relatively more Spanish-like than the L3 productions of the L2 subset are English-like.

[insert figure 6]

Discussion

Overall, the present study provides evidence that bilinguals are not consistently affected by the phonology of one language at first exposure to an L3. Rather, two group trends were found which reveal that a subset of participants primarily produced target like and L1-Spanish like relative VOT in L3 French. Oppositely, a larger subset of participants produced L3 relative VOT with heavy L2 influence. The L3 productions of the L2 subset were not practically equivalent to L2 productions however, suggesting that this subset was simultaneously influenced by both their languages, or by the L2 and the phonetic properties of the stimuli.

The results of the analyses were largely obtained with a statistically powered sample. A series of post-hoc power analyses of the difference between English and Spanish was 100 the full data set and the L1 and L2 subsets. Table 12 shows the statistical power of all possible comparisons using t-tests in each subset and language combination. Notably, the L1 subset has a power level of 18 in the Spanish-French comparison and the L2 subset has a power level of 45 for the English-French comparison. These low power levels suggest that the differences between these language pairings in each respective subset is not clear.

[insert table 12]

These results imply that the need for a sufficiently powered sample size was met, and that the results obtained here are likely not due to sampling error. At the same time, the L1 subset participants did not have a sufficient sample size to provide evidence for practical equivalence on a Test of Equivalence (TOST; Lakens, 2017). Given that these two trends of L1 or L2 influence on L3 productions, and their different degrees of influence, was not predicted, the present study did not have in mind potential subsets during the initial recruitment procedure. Future studies on absolute L3 beginners could benefit from even larger sample sizes ($n > 200$), coupled with a rich background questionnaire, in order to be able to robustly examine whether L1 groups are fully influenced by the L1, and to explore plausible correlations between linguistic background and choice of language in L3 productions.

The present data suggest that L3 learners at first exposure may ‘choose’ one language to primarily influence their L3, but that this primary influence could be simultaneously influenced by both the phonetics of the input or L1 influence. The present data cannot tease apart this difference, since the L1 of these participants, Spanish, and the L3, French, are both true-voicing languages. It is reasonable, however, to conclude that the presence of aspiration in L3 productions of these participants is due to L2 influence, since their L2, English, is an aspirating language, and these participants also aspirated in their L2 in the English EPT.

The present data are best explained by two L3 models, depending upon which subset is examined. The full data set and the L2 subset are best explained by the Linguistic Proximity Model. The hybrid values observed here could be explained by the LPM’s notion of L3 learning as co-activation of linguistic systems (Westergaard, Mitrofanova, Mykhaylyk, & Rodina, 2017). In the case of the L1 subset, the Typological Primacy Model (Rothman, 2011, 2013, 2015) and the LPM both may account for the results obtained, since it appears that L1 influence is rather strong, and the degree of potential L2 influence in the L1 subset

seems to be negligible. An issue with the L1 subset supporting the LPM is that it does not corroborate a specific prediction by the model and as such is not falsifiable, as scholars have argued (Puig-Mayenco et al., 2020). Unlike the TPM, the LPM makes fewer restrictive predictions of L3 behavior and allows for “Full Transfer Potential,” suggesting that either of the two previously acquired linguistic systems may fully transfer on a property-by-property basis, or simultaneously influence a single property. However, it is unclear why one language may influence L3 productions over another. In this case, clear L2 influence is seen in the complete dataset and within the L2 subset. This influence is not facilitative, since the L3 is Spanish-like in terms of VOT, and points to L2 status effects. However, it is unclear why the L1 subset group demonstrated target-like relative VOT in the L3. It is possible that these participants were simply more proficient mimics of the acoustic stimuli. On the other hand, it is also possible that these participants were meta-linguistically guided by knowledge that Spanish and French are historically similar languages. In order to evaluate whether the L1 subset was influenced more by the phonetics of the stimuli or L1 influence, a future study could involve a second L3 group which is given an aspirating L3 to shadow (such as German). Then, if participants were to aspirate in the L2 but not the L3, despite the fact that the stimuli would be aspirated, then evidence for L1 influence could be provided.

The present study did not lack limitations. Firstly, the participants completed a short background questionnaire, since individual differences were not included in the study’s original predictions. The lack of this information about linguistic background did not allow for an exploratory post-hoc analysis of the subset data. Second, the self-rated proficiency could have been corroborated with a standard proficiency measure such as the LexTALE in order to provide a more comprehensive background of language proficiency that could be used in further post-hoc exploratory analysis. Finally, the present data were exclusively collected online. Online data collection presents many challenges and limitations. Specifically, it was not possible to control for the environment of the participants in terms of ambient noise in their perception and production of the French stimuli. Variable audio

recording devices and speakers were used by participants. As a result, it cannot be ruled out that the present data could have been influenced by differences in audio quality produced by participants' equipment in both perception and production.

Conclusion

The present study examined the production of relative VOT of L3 French at first exposure by Mexican Spanish-English late bilinguals. The findings suggest that individual variability exists in the first productions of a third language in which speakers appear to be influenced by L1 and L2 phonology simultaneously. At the same time, the degree of language-specific influence was not equal; a subset of the full data set showed quite heavy L1 influence on L3 French production, where the L2 subset group produced hybrid values.

These findings replicated the findings in the literature which reported intermediate or hybrid VOT values in trilinguals and provides further evidence that both languages of a bilingual are active during L3 learning. At the same time, the degree of co-activation seems to vary widely. The present study may orient future research so that it may include groups which examine distinct L3 stimuli to better examine how the phonetics of the input interact with L2 status effects.

Additionally, the present study was the first to use a test of equivalence in L3 research. This statistical method may provide more precise insights as to whether the claims in the literature that L3 groups or within-subjects comparisons are actually equivalent. Future research should consider this method, rather than the use of a lack of a main effect or insignificant t-test as evidence for practical equivalence.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, 23(4), 459–484. <https://doi.org/10.1177/0267658307080557>
- Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Studies in Bilingualism* (Vol. 46, pp. 61–78). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.46.06bar>
- Brysbaert, M. (2020). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, 1–6. <https://doi.org/10.1017/S1366728920000437>
- Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for Language Acquisition: Comparing Adults' and Children's Patterns of Development in First, Second and Third Language Acquisition of Relative Clauses. *International Journal of Multilingualism*, 1(1), 3–16. <https://doi.org/10.1080/14790710408668175>
- Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, 7(1), 19–38. <https://doi.org/10.1080/14790710902972248>
- Lakens, D. (2017). Equivalence Tests: A Practical Primer for *t* Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, 8(4), 355–362.

<https://doi.org/10.1177/1948550617697177>

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, 20(3), 384–422.

<https://doi.org/10.1080/00437956.1964.11659830>

Llama, R., & Cardoso, W. (2018). Revisiting (Non-)Native Influence in VOT Production: Insights from Advanced L3 Spanish. *Languages*, 3(3), 30.

<https://doi.org/10.3390/languages3030030>

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, 7(1), 39–57. <https://doi.org/10.1080/14790710902972255>

Meisel, J. M. (1983). Transfer as a second-language strategy. *Language & Communication*, 3(1), 11–46. [https://doi.org/10.1016/0271-5309\(83\)90018-6](https://doi.org/10.1016/0271-5309(83)90018-6)

Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages*. Amsterdam: John Benjamins Publishing Company.

<https://doi.org/10.1075/sibil.40>

Plonsky, L., & Oswald, F. L. (2014). How Big Is “Big?” Interpreting Effect Sizes in L2 Research: Effect Sizes in L2 Research. *Language Learning*, 64(4), 878–912.

<https://doi.org/10.1111/lang.12079>

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research*, 36(1), 31–64. <https://doi.org/10.1177/0267658318809147>

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese.

IRAL - International Review of Applied Linguistics in Language Teaching,
48(2-3). <https://doi.org/10.1515/iral.2010.011>

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy:
The typological primacy model. *Second Language Research*, 27(1), 107–127.
<https://doi.org/10.1177/0267658310386439>

Rothman, J. (2013). Cognitive economy, non-redundancy and typological primacy in
L3 acquisition: Initial stages of L3 Romance and beyond. In S. Baauw, F.
Drijkoningen, L. Meroni, & M. Pinto (Eds.), *Romance Languages and Linguistic
Theory* (Vol. 5, pp. 217–248). Amsterdam: John Benjamins Publishing Company.
<https://doi.org/10.1075/rllt.5.11rot>

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological
Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and
proficiency considered. *Bilingualism: Language and Cognition*, 18(2), 179–190.
<https://doi.org/10.1017/S136672891300059X>

Slabakova, R. (2017). The scalpel model of third language acquisition. *International
Journal of Bilingualism*, 21(6), 651–665.
<https://doi.org/10.1177/1367006916655413>

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). EFFECTS OF AGE AND
SPEAKING RATE ON VOICE ONSET TIME: The Production of Voiceless
Stops by Near-Native L2 Speakers. *Studies in Second Language Acquisition*, 37(1),
71–100. <https://doi.org/10.1017/S0272263114000151>

Ullman, M. T. (2004). Contributions of memory circuits to language: The
declarative/procedural model. *Cognition*, 92(1-2), 231–270.
<https://doi.org/10.1016/j.cognition.2003.10.008>

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017).

Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, 21(6), 666–682.

<https://doi.org/10.1177/1367006916648859>

Williams, S., & Hammarberg, B. (1998). Language Switches in L3 Production:

Implications for a Polyglot Speaking Model. *Applied Linguistics*, 19(3), 295–333.

<https://doi.org/10.1093/applin/19.3.295>

Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, 7(1), 75–90. <https://doi.org/10.1080/14790710902972263>

Wrembel, M. (2014). VOT Patterns in the Acquisition of Third Language Phonology.

Proceedings of the International Symposium on the Acquisition of Second Language Speech, 5. Concordia Working Papers in Applied Linguistics,.